modeled for a time horizon of 35-40 years or for a lifetime to demonstrate cost effectiveness. **CONCLUSIONS:** This analysis shows the range, variability, and methods used for calculation of ICER values for these high budget impact drugs and provides lessons for executives and policy makers.

PRM40

SYSTEMATIC LITERATURE REVIEW TO EVALUATE AND CHARACTERIZE THE HEALTH ECONOMICS AND OUTCOMES RESEARCH STUDIES IN INDIA

Mishra D[1], Nair SR[2]

[1]*Pfizer Limited, Mumbai, Maharashtra, India,* [2]*Capita India Pvt. Ltd., Mumbai, Maharashtra, India*

OBJECTIVES: Health economics research is in its infancy in India. This systematic literature review aimed to identify, evaluate, and characterize the variety, quality, and intent of the health economics and outcomes studies being conducted in India. **METHODS:** Studies published in English language between 1999 and 2012 were retrieved from Embase and PubMed databases using relevant search strategies. Two researchers independently reviewed studies as per Cochrane methodology; information on type of research and outcomes were extracted. Quality of reporting was assessed for model-based health economic studies using a published 100-point Quality of Health Studies (QHES) instrument. Subjective assessment was used for the remaining studies. **RESULTS:** Of 546 studies screened, 132 studies were included in the review. The broad study categories were cost-effectiveness analyses ([CEA], 54 studies), cost analyses (19 studies), and burden of illness (18 studies). The outcomes evaluated in cost studies were direct and indirect costs, and incremental cost-effectiveness ratio (ICER), quality-adjusted life years (QALYs), and disability-adjusted life years (DALYs). Most studies were conducted from societal perspective. Direct medical costs assessed cost of medicines, monitoring costs, consultation and hospital charges along with non-medical costs (travel and food for patients and c are-givers). Loss of productivity and loss of income of patients and care-givers were identified as components of indirect cost. Overall, 33 studies assessed QoL, and WHO Quality of Life-BREF (WHOQOL-BREF) was the most commonly used instrument in these studies. Quality assessment for modeling studies showed that most studies were of high quality (mean [range] QHES score to be 75.5 [34-93]). **CONCLUSIONS:** This review identified various patterns of health economic studies in India. Majority of the CEA studies conducted in recent years were of high quality. Despite this, utilization of health care resources is inappropriate and economic evaluation needs relevance to the context of health care in India.

PRM41

A NOVEL APPROACH TO RANKING PARAMETER UNCERTAINTY IN ONE-WAY SENSITIVITY ANALYSIS: WHAT TORNADO DIAGRAMS ARE MISSING

Herring W, Mladsi DM, Miles L, Ronquest N

RTI Health Solutions, Research Triangle Park, NC, USA

OBJECTIVES: Economic evaluations of health technologies have long relied on one-way sensitivity analysis (SA) to examine the impact of parameter uncertainty on modeling outcomes. Traditionally, this impact has been measured and ranked based on absolute changes in the incremental cost-effectiveness ratio (ICER) across plausible parameter values and presented in a tornado diagram. This format does not adequately identify or prioritize parameters where the range of uncertainty causes the ICER to change quadrants in the cost-effectiveness (CE) plane. However, these quadrant changes, which represent fundamental changes to the CE conclusion, are arguably more meaningful than changes in the ICER within a quadrant. This research illustrates a novel approach to presenting one-way SA results that focuses on identifying parameters with the greatest potential to change the overall CE conclusion rather than narrowly focusing on changes to the ICER. **METHODS:** We developed a comprehensive algorithm for ranking the parameters varied in a one-way SA. Broadly, we first prioritize parameters with the potential to qualitatively change the CE conclusion and then rank parameters based on quantitative changes to modeling outcomes. Changes to the CE conclusion are identified based on quadrant changes in the CE plane, and parameters are categorized as having the potential to change the conclusion both positively and negatively, only positively, only negatively, or not at all. Within these categories, a secondary ranking based on costs and health outcomes is used. Furthermore, visualization techniques anchored in the CE plane help assess whether conclusion changes are due primarily to changes in health, changes in costs, or both. **RESULTS:** This research demonstrates that the conclusions-based ranking algorithm works in more general settings than the traditional tornado diagram format. **CONCLUSIONS:** The conclusions-based approach is a powerful method that provides a more complete picture of the impact of parameter uncertainty in economic evaluations.

PRM42

IDENTIFYING ACCURATE PATIENT-BASED HEALTH AND SOCIAL CARE COSTS OF OLDER PEOPLE FOR TRIAL-BASED ECONOMIC EVALUATION: IS IT REALLY WORTH IT?

Elliott RA[1], Franklin M[1], Berdunov V[1], Gkountouras G[1], Edmans J[1], Conroy S[2], Bradshaw L[1], Tanajewski L[1], Gladman J[1]

[1]*University of Nottingham, Nottingham, UK,* [2]*University of Leicester, Leicester, UK*

OBJECTIVES: Acute Medical Units (AMUs) identify individuals requiring in-patient care and those who can be discharged. However, readmission rates for older people in the year following AMU discharge are high. We aimed to identify patient-based health and social–care costs of a cohort of older (70+) people discharged home from AMU within 72 hours. **METHODS:** Although resource-use data from social and health care sectors in England is available electronically, there is little systems-linkage so data were obtained from each sector separately. Hospitalisation and social care data were collected retrospectively for 644 patients for three months post-AMU discharge using patient administration systems. In a subset (n=456), further approvals were gained for general practices, ambulance services, intermediate and mental health care. Of 118 general practices serving our cohort, data were obtained from 48 (250/456 participants). Seventeen were not covered by our approvals, and, despite an expert and dedicated team, 53 declined access or did not respond. **RESULTS:** We obtained data on hospitalisations for all participants, and "full" costs for 250 participants. Mean (95% CI, median, range) total cost for this subgroup was £2006 (1642-2470, 0, 0-23612). Secondary care constituted 76.2% costs. Contribution from other sectors was: primary care (10.9%), ambulance service (0.7%), intermediate care (0.1%), mental health care (2.1%) and social care (10.0%) The top 10% participants accounted for 50% of overall cost. **CONCLUSIONS:** This study was resource-intensive due to: complex approvals and access requirements; geographical dispersion of participants and data sources; different recording systems; varying data quality; different care definitions across geographical sites; with manual data extraction often required. Care of older people is generally moving from secondary care to primary health care and social care, so the costs of other sectors will increase in the future. There is clearly a need to improve access and system interoperability and streamline methods for obtaining these costs.

PRM43

THE LOWER AND UPPER LIMITS OF AN INCREMENTAL COST-EFFECTIVENESS RATIO ASSOCIATED WITH THE EFFICIENCY FRONTIER: A CASE OF HIV/AIDS PREVENTION AND TREATMENT

Kamae I[1], Araki D[2], Sugimoto T[3]

[1]*The University of Tokyo, Graduate School of Public Policy, Tokyo, Japan,* [2]*Meiji University of Integrative Medicine School of Nursing Science, Nantan-shi, Kyoto, Japan,* [3]*The University of Tokyo, Tokyo, Japan*

OBJECTIVES: To validate a theoretical method, using a case study of efficiency frontier analysis for HIV/AIDS prevention and treatment, for quantifying lower and upper limits of the ICER configurated as a slope of the connected lines on the efficiency frontier, whose research poster in theory was presented at the ISPOR Dublin 2007. **METHODS:** A validation study was conducted based on the evidence published on the Journal of Medical Economics as one of the first studies that identified cost-effectiveness variation in efficient frontiers for HIV/AIDS prevention and treatment documented in the database, Cost-Effectiveness Analysis Registry, at Tufts Medical Center in 2002-2007. **RESULTS:** Recognizing that there are two types of efficiency frontier represented with a monotonically increasing function of QALY on the Y-axis and Cost on the X-axis, and also each efficiency frontier is formulated by two regression models: log and square-root models, two types of the mathematical formulae of derivatives were obtained according to each regression model. The formula of derivatives can provide the slope of the tangent on the efficiency frontier curve, given an arbitrary value of cost (or QALY). Therefore, we developed the formula that can provide lower and upper slopes of the two tangents, given an arbitrary ICER slope which is represented by connecting two points on the efficiency frontier curve. The examples of calculations conducted were graphically illustrated according to each regression model. **CONCLUSIONS:** Through this validation study, we can confirm that an application of our method is theoretically and practically feasible to estimate the lower and upper limits of an ICER arbitrarily given on the efficiency frontier curve. This approach will provide us with more useful information on the question how we could interpret and justify the high value of ICER of new technologies such as molecular-targeted drugs.

PRM44

BIAS IN RELATIVE ACCURACY METRICS

Schmier JK[1], Cross PJ[1], Cantor RA[1], Lau EC[2], Steffey DL[2], Watson HN[2]

[1]*Exponent, Inc., Alexandria, VA, USA,* [2]*Exponent, Inc., Menlo Park, CA, USA*

OBJECTIVES: Reimbursement to beneficiaries and cost-effectiveness analyses depend on the availability of estimates of billed-charge amounts. In the United States, no single data source is universally accepted; rather, several vendors compile distributions of billed charges. Studies use these estimates interchangeably and there is no preferred metric for characterizing differences in conducting comparisons. Nonetheless, the conventional wisdom suggests that Medicare data tend to underestimate means and other values relative to commercial data. This paper investigates the statistical properties of three metrics used to characterize relative differences between two sources of values. For these metrics, the difference between the tested and reference values is the numerator; the denominators are the tested value (Metric #1), the reference value (#2), or the average of the two (#3). **METHODS:** Each metric is described; mathematical proofs and simulations demonstrate the types of bias that can be introduced. **RESULTS:** Two simulations of constructed distributions with identical means, one with small value differences and the other with large value differences, demonstrate that Metrics #1 and #2 would result in opposite inferences. Proofs demonstrate that the expected values of Metrics #1, #2, and #3 respectively, are negative, positive, and zero. The head to head comparison of two billed charge benchmarks finds bias present across all three averages (simple, weighted by claim count, and weighted by claim dollars), although it is most pervasive for the simple average. The sign of the average of Metric #1 is consistently negative, while that for Metric #2 is positive and Metric #3 varies. **CONCLUSIONS:** When many observations are aggregated to generate an overall average, the choice of metric affects results, often to the point where the choice of one methodology or another can generate diametrically opposite conclusions.

Results demonstrate the importance of sensitivity analyses and thoughtful consideration of the metric and benchmark selection.

PRM45

HEALTH RESOURCE UTILIZATION OF LUPUS NEPHRITIS PATIENTS – A COMPARISON OF RESULT ACROSS CASE IDENTIFICATION ALGORITHMS

Yeh WS, Clark DW, McCarty KM
Biogen Idec, Weston, MA, USA

OBJECTIVES: Lupus nephritis (LN) is a severe complication of systemic lupus erythematosus (SLE). While LN has no designated ICD-9 code, various approaches have been used to identify patients in administrative data. The objective was to compare health resource utilization of LN patients using different algorithms in a single data source. **METHODS:** This study used the Impact Database, a commercial insurance claims database. SLE patients were identified using ICD-9 code 710.0 from ≥2 outpatient or 1 inpatient claims from 01/2004 to 06/2011. SLE patients with LN were further identified under four different algorithms: (A) ≥1 renal diagnosis, (B) ≥2 renal diagnoses, (C) ≥3 renal diagnoses, and (D) ≥3 renal diagnoses plus ≥3 nephrologist visits. Health resource utilization and expenditure of outpatient/emergency department (ED) visits, hospitalizations, and prescriptions were examined for 12 months post index date of first renal diagnosis. **RESULTS:** A total of 93,957 patients were diagnosed with SLE. Among them, 24,357, 11,054, 8,895, and 6,307 cases had LN using algorithms A-D. LN cases identified by algorithms A-D had similar mean age (48.3, 46.7, 46.3, and 45.7 years) and gender distribution (85.2, 83.1, 82.7, and 81.8% females). LN patients from different algorithms also had similar annual frequency of outpatient visit (35.9, 41.0, 41.1, and 42.1), ED visit (1.4, 1.4, 1.4, and 1.5), hospitalization (0.8, 0.9, 0.9, and 0.9), and prescription (6.8, 7.3, 7.2, and 7.4). The annual medical expenditures were $33,176, $36,974, $36,241, and $38,883 for algorithms A, B, C, and D, respectively. **CONCLUSIONS:** Our results support that when studying health resource utilization, the results do not differ significantly based on the number of renal diagnoses codes. There is a difference in outcomes when requiring number of patients plus specialty subtype; however, in the case where specialty information is either unavailable or unreliable, using algorithms A-C proved equally reliable in an administrative claims database.

PRM46

ANALYSIS OF HEALTH CARE COSTS CONTAINING A LARGE PROPORTION OF $0 DATA USING TRADITIONAL AND ZERO-INFLATED GAMMA REGRESSION MODELS

Kern DM, Wasser T
HealthCore, Inc., Wilmington, DE, USA

OBJECTIVES: Compare traditional and zero-inflated gamma regression models for the analysis of health care costs in an administrative claims database. Gamma regression models are widely used for the analysis of health care cost data and are appropriate for analyzing mean costs within patients incurring >$0. However, little research is available for the appropriateness of gamma regression on data with a high proportion of $0 costs. **METHODS:** This study used the HealthCore Integrated Research Environment to analyze hospitalization costs within 3,049 bipolar patients initiating antipsychotic medications (Drug X=2,398 patients; Drug Y=651 patients). Results of a traditional gamma model and a zero-inflated gamma model were compared. The zero-inflated model used SAS procedure NLMIXED to perform logistic regression modelling the probability of having a hospitalization, gamma regression modelling mean costs within patients having a hospitalization, and combining the two models to analyze differences in overall mean cost. **RESULTS:** There were 18.3% Drug X patients with an inpatient hospitalization, compared with 12.1% of Drug Y. Mean costs were higher for Drug X within only patients having an event ($17,721 vs. $11,425) and including patients with no event ($3,237 vs. $1,387). Gamma regression found the difference in the overall population to be significant (mean-diff=$1,850, CI=[$1,463;$2,152], p<0.0001). The zero-inflated gamma regression model showed similar results (mean-diff=$1,850, CI=[$1,237;$2,464], p<0.0001). Consistent results between the models were also seen when adjusting for patient demographics, comorbidities, and prior medications: traditional gamma model (mean-diff=$2,173, CI=[$716;$1,297], p<0.0001), zero-inflated gamma model (mean-diff=$2,448, CI=[$597;$1,670], p<0.0001). **CONCLUSIONS:** When analyzing health care cost data containing >80% $0 costs, little difference was seen between traditional gamma regression and zero-inflated gamma regression models. The zero-inflated model uses complex coding, and requires advanced knowledge of statistical methods and SAS programming. Within data containing a large proportion of $0 costs, traditional gamma regression is appropriate for analyzing differences in mean costs.

PRM48

DIAGNOSTICS FOR CHECKING THE GAMMA DISTRIBUTION ASSUMPTION IN GENERALIZED LINEAR MODELS USED FOR MODELING HEALTH CARE COSTS

Juneau P[1], Pan KY[2], Curkendall SM[3]
[1]*Truven Health Analytics, Bethesda, MD, USA,* [2]*Truven Health Analytics, Cambridge, MA, USA,* [3]*Truven Health Analytics, Washington, DC, USA*

OBJECTIVES: 1) To study the relative performance of various graphical and analytical diagnostics used to assess the assumption of an underlying gamma distribution for a commonly used model in the examination of health care expenditures (GLMAGALL – a generalized linear model assuming a gamma distribution with a log link relating the mean costs to a corresponding set of predictors), and 2) To investigate each diagnostic's ability to discriminate between various assumed distributions (e.g., gamma versus log-normal) that may be potentially used in health care cost analysis. **METHODS:** Data sets will be created via Monte Carlo simulation of gamma (varying the shape and scale parameters) and other distributions, as well as, by altering the values for a small set of predictors. The results of the diagnostic study will be illustrated in a graphical and/or tabular format. **RESULTS:** Graphical diagnostics afford the analyst the ability to see subtle or dramatic departures from the model's distributional assumptions that might not be as obvious by using an analytical model that provides a single summary statistic. **CONCLUSIONS:** The performance of a diagnostic procedure to assess the presence of a gamma distribution in a cost model or its ability to discriminate between one distribution and another is important; however, other factors must be considered before an analyst makes his or her final choice. The ease of executing the technique, its relative clarity of interpretation, and availability in a software package (without having to resort to extensive programming beyond what is provided by a standard statistical package) must all be considered to ensure that model adequacy testing may be performed readily so that the choice of a distribution for an expenditure model may be considered sound.

RESEARCH ON METHODS – Databases & Management Methods

PRM50

USING TEXT MINING OF ELECTRONIC MEDICAL RECORDS TO IDENTIFY KRAS TESTING STATUS IN MCRC PATIENTS

Miller PJ[1], Walker MS[1], Landsman-Blumberg P[2], Cuyun Carter G[3]
[1]*ACORN Research, LLC, Memphis, TN, USA,* [2]*Truven Health Analytics, Bethesda, MD, USA,* [3]*Eli Lilly and Company, Indianapolis, IN, USA*

OBJECTIVES: To develop algorithms identifying if metastatic colorectal cancer (mCRC) patients were tested for KRAS (a tumor biomarker of EGFR-inhibitor response) using text documents (e.g., physician progress notes) within electronic medical records (EMR). **METHODS:** The sample consisted of 1,385 mCRC patients from the ACORN Data Warehouse. 300 patients were randomly selected for chart review and randomly assigned to training (n=150), validation (n=50), or testing datasets (n=100); 1,085 patients comprised a scoring dataset. Counts of terms in text-based content of patient EMRs were used to develop models predicting KRAS testing status. **RESULTS:** Several models were used to predict KRAS testing in the training sample. Decision tree (DT), random forest (RF), and adaptive boosting (AB) models performed best when applied to validation data not used in the earlier model development process. RF outperformed DT and AB. RF was the only model to produce a kappa ≥ 0.80 (within rounding) for both the validation and testing datasets. It also produced the highest kappa in the testing dataset (kappa=0.7994), as well as fewer false negatives. RF was used to score the remaining 1,085 patients. All patients predicted "tested" and a random sample of patients predicted "not tested" underwent chart review. The model correctly predicted KRAS "tested" 482/500 times (PPV=96.4%) and "not tested" 196/200 times (NPV=98.0%), kappa= 0.970. **CONCLUSIONS:** Text mining yielded highly accurate classification of KRAS testing status among mCRC patients. Review of the small number of misclassified cases of KRAS testing identified ways to improve the model's accuracy. These results may inform future research and reduce the need for labor-intensive and costly full chart review by human coders.

PRM51

THERE BUT FOR GRACE? A VALIDATED SCREENING TOOL FOR QUALITY OBSERVATIONAL STUDIES OF COMPARATIVE EFFECTIVENESS

Dreyer NA[1], Velentgas P[1], Westrich KD[2], Dubois RW[2]
[1]*Quintiles Outcome, Cambridge, MA, USA,* [2]*National Pharmaceutical Council, Washington, DC, USA*

OBJECTIVES: To be able to identify observational studies of good enough quality for decision support by validating a set of screening questions to qualify studies likely to produce reasonably accurate and unbiased estimates of comparative effectiveness (CE). **METHODS:** An 11-item checklist was developed through literature review and consultation with experts from ISPOR, ISPE, payer groups, private sector and academia. Item content covers four quality domains: comparability of subjects, information about the exposure or intervention, outcome measurement, and statistical analysis, which are metrics similar to those used in assessing observational study quality for systematic reviews. Checklist items were tested using studies of drugs, medical devices and medical procedures. We focused on research quality, not applicability to any decision. A fundamental challenge was to find a gold standard against which to test checklist items. 113 volunteers from 5 continents each rated >3 articles (N=280 assessments) from three validation sets of studies that 1) had quality assessments published in systematic reviews; 2) were assessed for quality by one of nine advisors from academic and payer groups; or 3) were assessed for quality by two of the nine advisors. **RESULTS:** Expert reviews uncovered an unsettling lack of agreement about what "good" looks like, especially in situations that lacked context, with 52% concordance (5 experts, 23 assessments.) The single best performing checklist item, data quality for the primary outcome(s), scored >= 0.67 for positive predictive value in 4 of 6 samples and > = 0.67 for negative predictive values in all 6 samples. Another high scoring question, sensitivity analyses, had a positive predictive value >= 0.69 for in all 6 samples. **CONCLUSIONS:** This quantitative study shows that many content items recommended by experts do not consistently distinguish high quality observational CE studies.

PRM52

ASSOCIATION BETWEEN CARDIOVASCULAR BIOMARKERS LEVELS AND CIGARETTE SMOKING AMONG CURRENT SMOKERS, PAST SMOKERS AND NON SMOKERS USING NHANES 2007-2010

Saxena K[1], Liang Q[2], Muhammad-Kah R[2], Sarkar M[2]
[1]*Virginia Commonwealth University, Richmond, VA, USA,* [2]*Altria Client Services Inc., Richmond, VA, USA*